Complex Systems Methods — 11. Power laws — an indicator of complexity?

Eckehard Olbrich

e.olbrich@gmx.de
http://personal-homepages.mis.mpg.de/olbrich/complex_systems.html

Potsdam WS 2007/08

Overview

Introduction

The ubiquity of power laws

- Observed in light from quasars, intensity of sunspots, flow of rivers such as the Nile, stock exchange price indices
- Distribution of wealth, city sizes, word frequencies
- Fluctuations in physiological systems (heart rate, ...

Why are power laws interesting?

- Heavy tails, higher probability of “untypical events”
- Power law distributions are self-similar
- Occur in fractals
- Might be related to some optimality principle

Power laws for discrete and continuous variables

- A quantity x obeys a power lay if it is drawn from a probability distribution
$$p(x) \propto x^{-\alpha}$$
- If x is a continuous variable, we have a probability density function (PDF)
$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$
- If $x \in \mathcal{N}$ $x \geq x_{min}$, the probability of a certain value x is given by
$$p(x) = \frac{1}{\zeta(\alpha, x_{min})} x^{-\alpha}$$
with the generalized zeta function
$$\zeta(\alpha, x_{min}) = \sum_{n=0}^{\infty} (n + x_{min})^{-\alpha} \ .$$

Cumulative distribution function and rank/frequency plots

- Best way to plot power laws is to plot the cumulative distribution function (CDF) $P(x)$ — the probability to observe a x' that has a value that is equal or larger than x — no binning necessary, lower statistical fluctuations
- In the continuous case

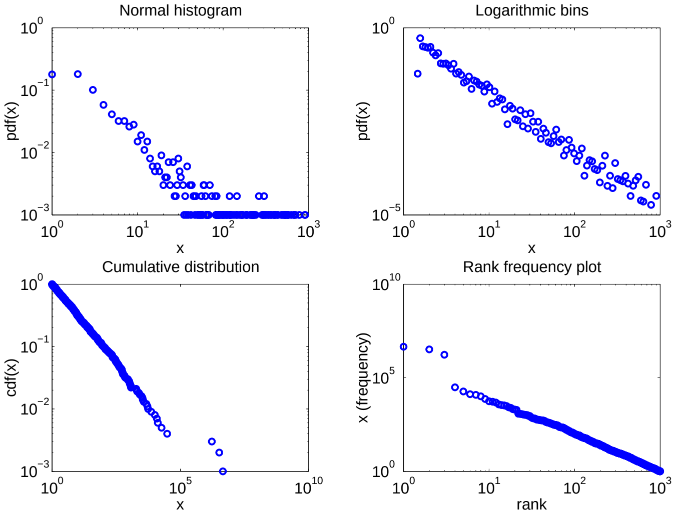
$$P(x) = \int_x^\infty p(x')dx = \left(\frac{x}{x_{min}}\right)^{-\alpha+1}$$

- In the discrete case

$$P(x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{min})}$$

- Plotting the number of x''s which are equal or larger than x with x being the frequency of occurrence gives the rank/frequency plot, which corresponds to the (unnormalized) cumulative distribution function

Plotting power laws



Zipf's law and its relatives

- Zipfs law: in natural langage the frequency of words is inversely proportional to its rank in the frequency table
- George Kingsley Zipf (1902-1950), American linguist and philologist who studied statistical occurrences in different languages.
- Vilfredo Federico Damaso Pareto (1848 - 1923) was a French-Italian sociologist, economist and philosopher.
- Pareto noticed that 80% of Italy's wealth was owned by 20% of the population.
- Pareto distribution: Number of people with an income larger than x.
- Pareto index is the exponent of the cumulative distributions of incomes

Fitting the exponent

- Usual way: making a histogram and fitting a straight line in a log-log plot
- Problems: The errors are not gaussian and have not the same variance
- Maximum likelihood estimate of the slope $\hat{\alpha}$ such that $p(\alpha|x_1, \ldots, x_n)$ becomes maximal.
- Probability density

$$p(x) = Cx^{-\alpha} = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

- Probability to observe a sample of data $\{x_1, \ldots, x_n\}$ given α

$$p(x_1, \ldots, x_n|\alpha) = \prod_{i=1}^{n} p(x_i) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha}$$

Fitting the exponent

- Probability to observe a sample of data $\{x_1, \dots, x_n\}$ given α

$$p(x_1, \dots, x_n|\alpha) = \prod_{i=1}^{n} p(x_i) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha}$$

- Bayes rule

$$p(\alpha|x_1, \dots, x_n) = p(x|\alpha) \frac{p(\alpha)}{p(x_1, \dots, x_n)} \ .$$

- Assuming an uniform prior this leads to maximizing the log likelihood

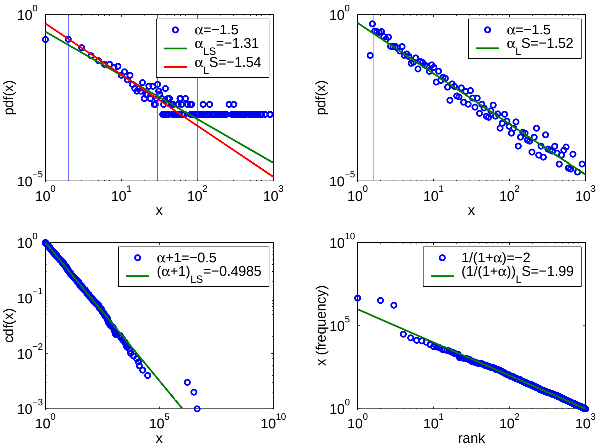
$$\begin{aligned} L &= \ln p(x_1, \dots, x_n|\alpha) \\ &= n \ln(\alpha - 1) - n \ln x_{min} - \alpha \sum_{i=1}^{n} \ln \frac{x_i}{x_{min}} \end{aligned}$$

- Setting $\partial L / \partial \alpha = 0$, one finds

$$\boxed{\hat{\alpha} = 1 + n \left(\sum_i \ln \frac{x_i}{x_{min}} \right)^{-1}}$$

Comparism with least square fit in the log-log plot

Maximum likelihood estimate: $\hat{\alpha}_{MLE} = 1.5003$ using all data points



Maximum likelihood estimator of α

$$\boxed{\hat{\alpha} = 1 + n\left(\sum_i \ln \frac{x_i}{x_{min}}\right)^{-1}}$$

with standard error

$$\sigma_{\hat{\alpha}} = \frac{\hat{\alpha} - 1}{\sqrt{n}} + O(1/n)$$

More precise results in Clauset at al. (2007)

$$\begin{aligned} \langle \hat{\alpha} \rangle &= \alpha \frac{n}{n-1} - \frac{1}{n-1} \\ \sigma_{\hat{\alpha}} &= (\alpha - 1)\frac{n}{(n-1)\sqrt{n-2}} \end{aligned}$$

Estimating the lower bound x_{min}

- Often $\hat{x}_m in$ has been chosen by visual inspection of a plot of the probability density function or the cumulative density function
- Increasing x_{min} means unsing less data for the fit, thus the likelihood cannot be compared directly
- Bayes'sches information criterion (BIC) having n points with $x_i > x_{min}$

$$\ln P(x|x_{min}) \simeq L - \frac{1}{2}(x_{min} + 1) \ln n \,.$$

- BIC underestimates x_{min} if there is a simple (but not a power) law also for a region below x_{min}.
- Clauset et al. propose to minimize the difference between the empirical distribution and the power law disribution using the Kolmogorov-Smirnov (KS) statistic

$$D = \max_{x \geq x_{min}} |S(x) - P(x, \hat{\alpha}, x_{min})|$$

with $S(x)$ being the CDF of the data for the observations with value at least x_{min}.

Testing the power law hypothesis

Two aspects:

1. Is the power law hypothesis compatible with the data and not ruled out?
2. Are there other better explanations of the data?

1. Testing the power law: Clauset et al. propose a Monte Carlo procedure, i.e. (1) generating a large number of sysnthetic data sets drawn from the power-law distribution that best fits the observed data, (2) fit each one individually to the power-law model, (3) calculate the KS statistic for each one relative to its own best-fit model, (4) count the fraction that the resulting statistic is larger than the value D observed for the true data. This fraction is the *p-value*. (5) If the p-value is sufficiently small, the power law hypothesis can be ruled out.

Testing the power law hypothesis

Two aspects:

1. Is the power law hypothesis compatible with the data and not ruled out?
2. Are there other better explanations of the data?

2. Testing against alternative explanations: Calculating the log likelihood ratio

$$\begin{aligned}\mathcal{R} &= \ln \frac{L_1}{L_2} = \ln \prod_{i=1}^{n} \frac{p_1(x_i)}{p_2(x_i)} \\ &= \sum_{i=1}^{n} (\ln p_1(x_i) - \ln p_2(x_i)) = \sum_{i=1}^{n} \left(l_i^{(1)} - l_i^{(2)} \right)\end{aligned}$$

If $\mathcal{R}$ is significantly larger than 0 than p_1 is a better fit than p_2.

Transformation method

If x is distributed according to $p(x)$ and y is a function of x, $y = f(x)$, then we have

$$q(y)dy = p(x)dx$$

and therefore

$$q(y) = p(x)\left(\frac{df}{dx}\right)^{-1}_{x=f^{-1}(y)}$$

If x are uniformly distributed random numbers with $(0 \leq x < 1)$, we can ask for a function f, such that $y = f(x)$ is distributed according to a power law

$$q(y) = \frac{\alpha - 1}{y_{min}}\left(\frac{y}{y_{min}}\right)^{-\alpha} .$$

Thus

$$\frac{dy}{dx} = \frac{y_{min}}{\alpha - 1}\left(\frac{y}{y_{min}}\right)^{\alpha}$$

Transformation method

If x is distributed according to $p(x)$ and y is a function of x, $y = f(x)$, then we have

$$q(y)dy = p(x)dx$$

and therefore

$$q(y) = p(x) \left(\frac{df}{dx} \right)^{-1}_{x=f^{-1}(y)}$$

If x are uniformly distributed random numbers with $(0 \leq x < 1)$, we can ask for a function f, such that $y = f(x)$ is distributed according to a power law

$$q(y) = \frac{\alpha - 1}{y_{min}} \left(\frac{y}{y_{min}} \right)^{-\alpha} .$$

Thus

$$\frac{dy}{dx} = \frac{y_{min}^{1-\alpha}}{\alpha - 1} y^{\alpha}$$

Transformation method

$$\frac{dy}{dx} = \frac{y_{min}^{1-\alpha}}{\alpha - 1} y^{\alpha}$$

Integrating

$$\int_{y_{min}}^{y} \frac{dy'}{y'^{\alpha}} = \frac{y_{min}^{1-\alpha}}{\alpha - 1} \int_0^x dx'$$

gives

$$y = y_{min}(1-x)^{-\frac{1}{\alpha-1}}$$

Combinations of exponentials

A variant of the preceeding is the following:

- Suppose y has an exponential distribution $p(y) \propto e^{ay}$
- E.g. intervals between events occurring with a constant rate are exponentially distributed
- A second quantity $x = e^{by}$, then

$$p(x) = p(y)\frac{dy}{dx} \propto e^{ay}(1/b)e^{-by} = \frac{1}{b}x^{-1+\frac{a}{b}}$$

Zipfs law for random texts

This can explain Zipf's like behaviour of words in random texts:

- m letters, randomly typed, probability q_s of pressing the space bar, probability for any letter

$$q_l = (1 - q_s)/m$$

- Frequency x of of a particular word with y letters followed by a space

$$x = \left(\frac{1-q_s}{m}\right)^y q_s \propto e^{by} \quad \text{with} \quad b = \ln(1-q_s) - \ln m.$$

- Number of distinct possible words with length y goes up exponentially as $p(y) \propto m^y = e^{ay}$ with $a = \ln m$.
- Thus $p(x) \propto x^{-\alpha}$ with $\alpha = 1 - \frac{a}{b}$

$$\alpha = 1 - \frac{\ln m}{\ln(1-q_s) - \ln m} = \frac{\ln(1-q_s) - 2\ln m}{\ln(1-q_s) - \ln m} \approx 2 \quad \text{for large } m$$

Preferential attachment — Yule process

- According to Newman "one of the most convincing and widely applicable mechanisms for generating power laws"
- First reported for the size distribution of biological taxa by Willis and Yule 1922
- "The rich become richer"
- The model:
  - $n \ldots$ time steps, generations, at each step a new entity (city, species, ...) is formed. Thus the total number of entities is n.
  - $p_{k,n}$ fraction of entities of size k at step n. The total number of entities of size k is $np_{k,n}$.
  - At each step also m entities increase in size by 1 with a probability proportional to their size $k_i / \sum_i k_i$. Note that $\sum_i k_i = n(m+1)$.
  - Power law:

$$p_k \propto k^{-(2+\frac{1}{m})}$$

Preferential attachment — Yule process

- Expected increas of entities of size k is then

$$\frac{mk}{n(m+1)} np_{k,n} = \frac{m}{m+1} kp_{k,n}$$

- Master equation for the new number $(n+1)p_{k,n+1}$ of entities of size k:

$$(n+1)p_{k,n+1} = np_{k,n} + \frac{m}{m+1}\left[(k-1)p_{k-1,n} - kp_{k,n}\right]$$

  with the exception of entities of size 1

$$(n+1)p_{1,n+1} = np_{1,n} + 1 - \frac{m}{m+1} p_{1,n}$$

- What is the distribution in the limit $n \to \infty$?

$$p_1 = \frac{m+1}{2m+1} \qquad p_k = \frac{m}{m+1}\left[(k-1)p_{k-1} - kp_k\right]$$

Preferential attachment — Yule process

- What is the distribution in the limit $n \to \infty$?

$$p_1 = \frac{m+1}{2m+1} \qquad p_k = \frac{m}{m+1}\left[(k-1)p_{k-1} - kp_k\right]$$

$$\text{leading to} \qquad p_k = \frac{k-1}{k+1+1/m}p_{k-1}$$

- By iteration one gets

$$p_k = \frac{(k-1)(k-2)\dots 1}{(k+1+1/m)(k+1/m)\dots(3+1/m)}p_1$$

which can be expressed using the gamma function
$\Gamma(a) = (a-1)\Gamma(a-1)$, $\Gamma(1) = 1$

$$\begin{aligned} p_k &= (1+1/m)\frac{\Gamma(k)\Gamma(2+1/m)}{\Gamma(k+2+1/m)} \\ &= (1+1/m)\mathrm{B}(k, 2+1/m) \end{aligned}$$

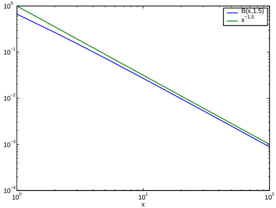
Preferential attachment — Yule process

- What is the distribution in the limit $n \to \infty$?

$$p_1 = \frac{m+1}{2m+1} \qquad p_k = \frac{m}{m+1}\left[(k-1)p_{k-1} - kp_k\right]$$

$$\begin{aligned} p_k &= (1+1/m)\frac{\Gamma(k)\Gamma(2+1/m)}{\Gamma(k+2+1/m)} \\ &= (1+1/m)\mathrm{B}(k, 2+1/m) \end{aligned}$$

- $B(a,b)$ has a power law tail a^{-b}, thus $p_k \propto k^{-(2+\frac{1}{m})}$



Preferential attachment — Yule process

- What is the distribution in the limit $n \to \infty$?

$$p_1 = \frac{m+1}{2m+1} \qquad p_k = \frac{m}{m+1}\left[(k-1)p_{k-1} - kp_k\right]$$

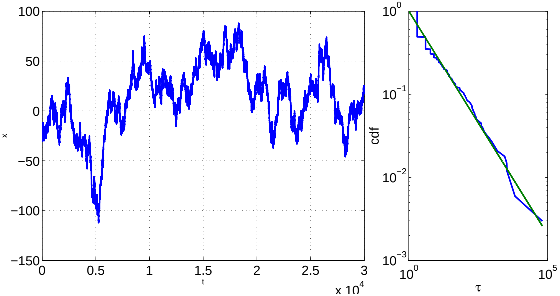
- $B(a, b)$ has a power law tail a^{-b}, thus $p_k \propto k^{-(2+\frac{1}{m})}$
- Generalizations: starting size k_0 instead of 1
- Attachment probability proportioanl to $k + c$ instead of k in order to handle also the case of $k_0 = 0$ (e.g. citations)
- Exponent

$$\alpha = 2 + \frac{k_0 + c}{m}$$

- Most widely accepted explanation for citations, city populations and personal income

- Random walks: Distribution of time intervals between zero crossings — "Gamblers ruin"

$$p(\tau) \propto \tau^{-3/2}$$



$$\hat{\alpha}_{MLE} = 1.5375$$

Multiplicative processes

- Multiplicative random processes: the logarithm of the product is the sum of the logarithms ⇒ the logarithm is normal distributed, the product itself is log-normal distributed

$$
\begin{aligned}
p(\ln x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right) \\
p(x) &= p(\ln x)\frac{d \ln x}{dx} = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right)
\end{aligned}
$$

The log-normal distribution is not a power law, but it can look like a power law in the log-log plot

$$
\ln p(x) = -\ln x - \frac{(\ln x - \mu)^2}{2\sigma^2} = -\frac{(\ln x)^2}{2\sigma^2} + \left(\frac{\mu}{\sigma^2} - 1 \right) \ln x - \frac{\mu^2}{2\sigma^2}
$$

Multiplicative processes

- Multiplicative time series model

$$x_{t+1} = a(t)x(t)$$

with $a(t)$ randomly drawn from some distribution gives still a log-nomal distribution.

- Kesten multiplicative process

$$x_{t+1} = a(t)x(t) + b(t)$$

produces a CDF for x with tails $P(x) \simeq \frac{c}{x^{\mu}}$ under the following conditions (see Sornette): a and b are i.i.d. real-valued random variables and there exists a μ such that

1. $0 < \langle |b|^{\mu} \rangle < +\infty$
2. $\langle |a|^{\mu} \rangle = 1$
3. $\langle |a|^{\mu} \ln |a| \rangle < +\infty$

This can be considered as a generalization of the random walk $a = 1$.

Criticality revisited

- Main feature of the critical state is the divergence of the correlation length (and/or time), which means that there is no typical length (time) scale in the system and the system has to be scale free.
- In equlibrium phase transitions power law dependencies between control and order parameters were a consequence of the self-similarity of the critical state.
- By using the same argument (see e.g. Newman) one can show, that e.g. size distribution of clusters at the percolation transition has to obey a power law.
- Models of SOC (also the game of life) can be related to directed percolation or Reggeon field theory, respectively, a certain kind of branching process.

Power laws and complexity

- The observation of a power law alone is not sufficient to call a system "critical" or even "complex".
- There are plenty of trivial and less trivial mechanims to generate power laws.
- But, power laws are interesting: A lot of (in some sense) "complex systems" show power laws. It might, however, not necessarily be related to their "complexity".